Fundamental Limits of Cooperation

Angel Lozano, Robert W. Heath Jr. and Jeffrey G. Andrews

Abstract

Cooperation is viewed as a key ingredient for interference management in wireless systems. This paper shows that cooperation has fundamental limitations. The main result is that even full cooperation between transmitters cannot in general change an interference-limited network to a noise-limited network. The key idea is that there exists a spectral efficiency upper bound that is independent of the transmit power. First, a spectral efficiency upper bound is established for systems that rely on pilot-assisted channel estimation; in this framework, cooperation is shown to be possible only within clusters of limited size, which are subject to out-of-cluster interference whose power scales with that of the in-cluster signals. Second, an upper bound is also shown to exist when cooperation is through noncoherent communication; thus, the spectral efficiency limitation is not a by-product of the reliance on pilot-assisted channel estimation. Consequently, existing literature that routinely assumes the high-power spectral efficiency scales with the log of the transmit power provides only a partial characterization. The complete characterization proposed in this paper subdivides the high-power regime into a *degrees-of-freedom regime*, where the scaling with the log of the transmit power holds approximately, and a *saturation regime*, where the spectral efficiency hits a ceiling that is independent of the power. Using a cellular system as an example, it is demonstrated that the spectral efficiency saturates at power levels of operational relevance.

I. INTRODUCTION

Wireless networks with many uncoordinated transmitters and receivers using the same spectrum are interference-limited, which means that increasing the transmit power of each node does not improve the spectral efficiency once the power is sufficiently high. In cellular systems, this results in a large fraction of users having low signal-to-interference-and-noise ratio (SINR) regardless of the transmit powers, unless bandwidth-wasting methods such as frequency reuse are implemented to relieve edge

users. Similar effects can be observed in other networks: in WiFi, inefficient contention-based MAC protocols protect receivers from interference by silencing nearby transmitters.

It has been persuasively argued in a by-now vast literature (e.g. [1]–[3]) that this limitation is not fundamental, but rather an artifact of each transmit-receiver pair communicating autonomously rather than cooperatively. If the various nodes could cooperate, the logic goes, the corresponding interference channel could be converted to a (perhaps very large) broadcast channel—for the downlink— or multiple access channel—for the uplink—with all the transmitters (resp. receivers) jointly encoding (resp. decoding). In the cellular context, it seems from this line of work that an arbitrary number of base stations (BSs) could cooperate to achieve enormous spectral efficiency gains over the lone-BS model [4], with the only limitation being the amount of coordination that is practical [5].

This is currently a problem of considerable theoretical and commercial interest. Incomplete but significant cooperation between BSs has been extensively attempted by industry and is still ongoing, broadly under the current moniker of "coordinated multipoint" (CoMP). The bit rate and latency of the backhaul links have limited the benefits of such cooperation thus far, with disappointing gains typically not exceeding 30% [6]–[8]. In fact, one respected group has even experienced a net loss from cooperative techniques when the various over-the-air overheads are accounted for [9]. Are these widespread observations a byproduct of current technology limitations that could be overcome with better/more backhaul sharing (e.g., over dedicated fiberoptic control channels), improved feedback and overhead techniques, and/or better encoding and decoding methods? Or is there a fundamental limitation lurking beneath the surface, one that is independent of the particular technology?

A. Background and Status Quo

There are currently about 3 million macrocellular BSs worldwide and that number is expected to reach 50 million by around 2015, once small cells (pico and femtocells) are incorporated [10]. In a typical urban area, thousands to tens-of-thousands of BSs occupy the same spectrum. Because of the ubiquity and commercial importance of cellular networks, plus the clear scope for cooperation in them, and in order to be concrete and clear, we focus the discussion around BS cooperation. However, the subsequent models and main results of this paper apply to any set of transmitters and receivers and their scope is not limited to cellular networks.

Evaluating the performance of cellular networks is a complex task. Large-scale computer simulations are ultimately necessary to verify the performance of any specific technique, but they are hardly the best way to devise and probe ideas, build intuition, and glean insights. Most designs in cellular systems are therefore incubated in much simpler settings that represent a fragment of a system and only eventually

are they transplanted and trialed at the level of an entire system. A typical such controlled setting (at a given time epoch) is given by the following relationship.

Relationship 1 *The observation at receiver n is*

$$Y_n = \sum_{k=1}^{K} H_{nk} \sqrt{P} X_k + Z_n \qquad\qquad n = 1, \ldots, N \qquad (1)$$

where K and N are the number of transmitters and receivers, respectively, X_k is the signal generated by transmitter k, normalized such that P is its power, and H_{nk} is the channel from transmitter k to receiver n. Power differences at the K transmitters can be simply absorbed into the channel coefficients or the noise variances. The term Z_n is the noise at receiver n, typically white and Gaussian, with some normalized variance.

The signals and channel coefficients in Relationship 1 can be scalars, or else properly dimensioned vectors and matrices to accommodate MIMO (multiple-input multiple-output) techniques. For the sake of exposition, unless otherwise stated, we henceforth consider single-antenna transmitters and receivers. Thus, the elements of Relationship 1 are scalars.

Relationship 1 subsumes, in effect, most basic information-theoretic channel settings.

- Single-user channel if $K = N = 1$.
- Multiaccess channel (MAC) if $K > 1$ and $N = 1$.
- Broadcast channel (BC) if $K = 1$ and $N > 1$.
- Interference channel (IC) if $K = N$ with $K, N > 1$.

Combinations of these basic settings are also possible. Relationship 1 has led to characterizations of the single-user, MAC and BC capacity, as well as asymptotic notions like the DMT (diversity-multiplexing tradeoff) [11], and practical techniques such as multiuser MIMO.

Relationship 1 has been used to study the IC both generically and in the context of cellular systems. Despite the fact that the capacity of the IC is yet to be determined, Relationship 1 has led to advances in the understanding of the IC.

- The definition of relevant quantities such as the number of DoF (degrees of freedom) [12].
- The development of centralized cooperative schemes such as Network MIMO [13].
- The genesis of distributed cooperative solutions such as IA (interference alignment) [14], Max-SINR (maximum signal-to-interference-plus-noise ratio) [15], [16] and other forms of cooperative interference management [17].

Most development on the IC takes place in the high-power regime, which is where the nature of the IC comes to the fore. For any setting conforming to Relationship 1, the high-power behavior is as illustrated
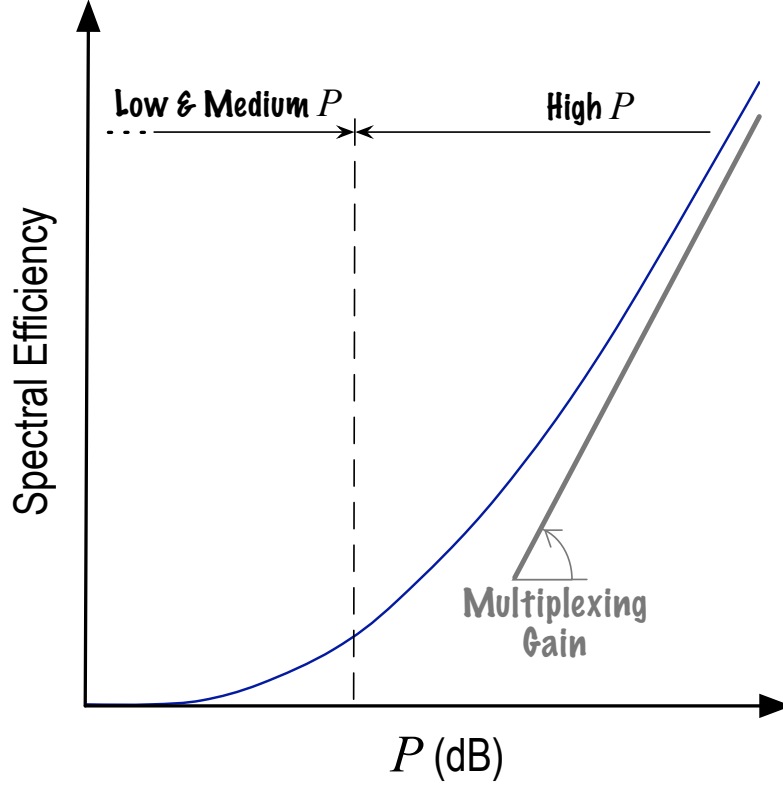
Fig. 1. Spectral efficiency as function of P (in dB) according to Relationship 1.

in Fig. 1. With the channel coefficients $\{H_{nk}\}$ known by the corresponding receivers and possibly also the transmitters, the spectral efficiency achieved by a given user grows, for $P \to \infty$, linearly with $\log(P)$ with a slope given by the so-called multiplexing gain. The multiplexing gain cannot exceed the number of DoF, and thus cooperative techniques such as Network MIMO or IA aim at maximizing the number of DoF.

The IC developments spawned by Relationship 1 have spurred an extensive amount of publications that invariably promise very large gains in spectral efficiency through cooperation (cf. [18]–[20] and references therein). However, as noted at the outset, in subsequent system-level simulations these gains do not seem to materialize: DoF maximizations attained through IA translate only to marginal gains or even outright losses in spectral efficiency [21], [22]. For example, [23], showed that a 300% gain from using IA in an isolated 3-user setting shrinks to a mere 28% gain when the same exact IA scheme is applied to a 19-cell system. (The conditions are otherwise identical: transmitters and receivers have 4 antennas and the signal-to-noise ratio is 20 dB.) These serial many-fold discrepancies point to a disconnect, to some fundamental way in which a fragment of a cellular system is not properly modeled by Relationship 1.

B. Modeling a Cluster Within a System

An obvious problem with Relationship 1 is that there is a cutoff of K (possibly cooperating) transmitters, and all other interference is ignored. The K cooperating transmitters are typically geographical neighbors, and are referred to as a *cluster*. It is usually assumed that any interference from outside the cluster can be lumped into the Z_n noise terms. It cannot, because Z_n has a fixed variance that does not depend on P whereas the external interference power is proportional to P. This can be illustrated by generalizing Relationship 1 as follows, to more accurately describe a wireless network.

Relationship 2 (Proposed System Model)

The observation at receiver n is

$$Y_n = \sum_{k=1}^{K} H_{nk}\sqrt{P}X_k + \sum_{k=K+1}^{\tilde{K}} H_{nk}\sqrt{P}X_k + Z_n \quad n = 1, \ldots, N \tag{2}$$

where $\tilde{K}$ and $\tilde{N}$ are the total numbers of transmitters and receivers in the system while K and N are the ones cooperating. Defining $Z_n' = \sum_{k=K+1}^{\tilde{K}} H_{nk}X_k$ as the out-of-cluster interference at receiver n,

$$Y_n = \sum_{k=1}^{K} H_{nk}\sqrt{P}X_k + \sqrt{P}Z_n' + Z_n \quad n = 1, \ldots, N \tag{3}$$

Thus, as $P \to \infty$, the out-of-cluster interference does as well if the total network size is greater than K. This gives rise to a *saturation regime* as shown in Fig. 2, where further increasing the power above some value P_{sat} does not noticeably increase the spectral efficiency because the external interference Z' is larger than the noise Z. Previous work had observed that interference power scales with P [24], [25], but the representation therein differs from Relationship 2 in two important ways. First, in [24], [25], the channel gains are deterministic and thereby known perfectly—at no cost—by all transmitters and receivers. In Relationship 2, in contrast, the channel gains outside the cluster of interest cannot be known (that is in fact what will come to define the boundaries of a cluster) and those within the cluster have to be learned, explicitly or not. Second, in [24], [25], a specific ratio of noise to interference is required, whereas in Relationship 2 that can be arbitrary. The assumptions in [24], [25] leads to the conclusion that the spectral efficiency still scales indefinitely with $\log(P)$, only with a modified DoF notion dubbed *Generalized DoF*. Relationship 2, in full generality, leads to a very different conclusion.

C. Summary of Contributions

This paper establishes that the correct representation of a cluster within a large wireless system is given by Relationship 2 and not Relationship 1, and that cooperation behaves very differently in the two models. A possible objection to this claimed distinction could be that the high-power saturation
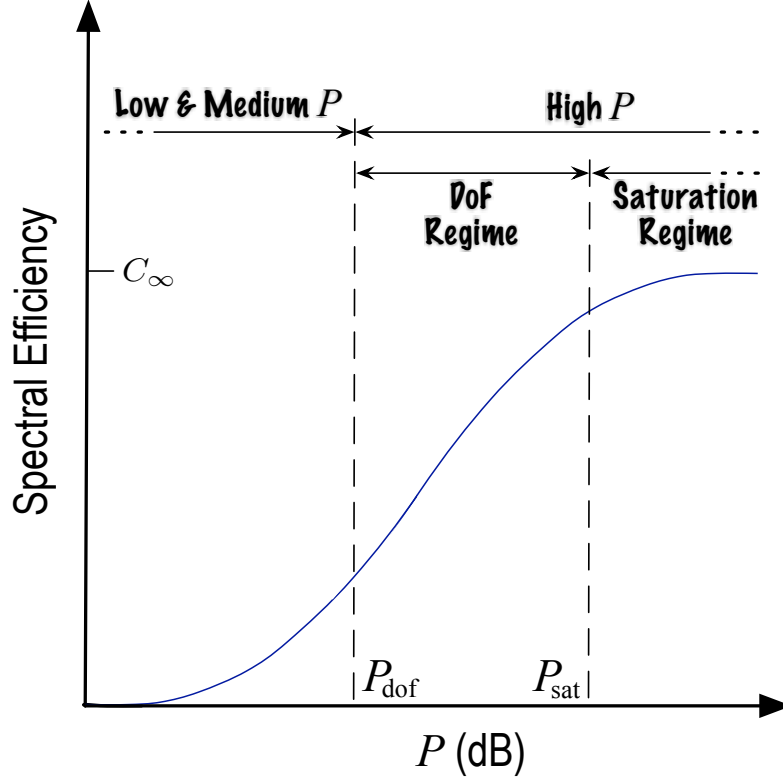
Fig. 2. Spectral efficiency as function of P (in dB) according to Relationship 2.

only occurs because the cluster size K, i.e., the number of cooperating transmitters, is too small. If K were made large enough, wouldn't the saturation dissappear since interfering transmitters would be converted to cooperating ones? The answer this paper establishes is negative. Saturation of the spectral efficiency is unavoidable, and cooperation can at best push it into higher power levels. This limitation holds in wireless networks with any degree of dynamics', i.e., channels that change over time or users that enter and leave the system. Virtually all wireless networks exhibit such dynamics. Although the derived limits may or may not be observed in practice today, they impose a hard ceiling on the gains from cooperative communication that can be achieved in the future. In particular, information theoretic fundamentals lead to the conclusion that it is not possible to convert a sufficiently large interference network to a BC or MAC regardless of the amount of cooperation.

The points established descend from the model detailed in Section II) and a fairly straightforward analysis of it for pilot-assisted channel estimation (in Section III) as well as a more fundamental analysis (in Section IV). The main result of the paper is Proposition 1, which establishes that there is an inescapable spectral efficiency upper bound in large cooperative networks. A large system (asymptotic) expression for this bound is given in Proposition 2. The main observations from the analysis and

accompanying examples are as follows.

1) In large systems, the existence of out-of-cluster interference (whose power scales with P) is inevitable. This holds regardless of whether channel coefficients are explicitly estimated or not, and regardless of the size of the clusters within the system.

2) As a result of the out-of-custer interference, a cluster within a cellular system is described by Relationship 2 rather than Relationship 1. The performance of any cooperative technique then no longer resembles Fig. 1 but rather Fig. 2.

3) For high P, two distinct regimes can be identified: a *DoF regime* where the noise dominates over the out-of-cluster interference and a *saturation regime* where the two become comparable and the spectral efficiency chokes. The notion of DoF is only meaningful in the DoF regime.

4) Two new quantities of interest emerge for the saturation regime. The value P_{sat} is the point where the transition to saturation is said to occur. The resulting limiting spectral efficiency is denoted by C_∞. These new quantities depend on the system topology, the channel propagation laws, and the degree of user mobility. In most cases, the transition to the saturation regime takes place well within the range of operational interest and thus studies conducted using Relationship 1 rather than (3) are bound to be misleading in terms of system-level performance.

5) The saturation takes place both with pilot-assistedcommunication and with noncoherent communication. Therefore the saturation does not occur simply because of explicit channel estimation or because of pilot overhead in the face of finite coherence times. Nor can improved channel estimation solve the problem.

II. DETAILED SYSTEM MODEL

This section is devoted to refining Relationship 2 and to describing the models used to embody it in the remainder of the paper.

A. Large-Scale Modeling

Let G_{nk} be the *average* channel power gain (associated with distance decay, shadowing, building penetration losses, and antenna patterns) between transmitter k and receiver n and let us also define the normalized channel power gains

$$g_{nk} = \frac{G_{nk}}{\sum_{\kappa=1}^{K} G_{n\kappa}} \tag{4}$$

such that, for every n,

$$\sum_{k=1}^{K} g_{nk} = 1. \tag{5}$$

Thus, g_{nk} signifies the share of receiver n's signal power that corresponds to transmitter k and the set $\{g_{nk}\}$ completely and compactly characterizes the relative average gains among all the nodes in the network. We term the set $\{g_{nk}\}$ the *geometry profile*.

We can further absorb the various normalizations for each receiver n into a signal-to-noise ratio (SNR_n) that scales with P and a signal-to-(out-of-cluster)-interference ratio (SIR_n) that does not scale with P. Specifically,

$$\mathsf{SNR}_n = \frac{\sum_{k=1}^{K} G_{nk} P}{N_0 B} \tag{6}$$

$$\mathsf{SIR}_n = \frac{\sum_{k=1}^{K} G_{nk}}{\sum_{k=K+1}^{\tilde{K}} G_{nk}} \tag{7}$$

where N_0 is the noise spectral density and B the bandwidth.

Using the definitions in (5), (6) and (7), we can rewrite (3) as

$$y_n = \sqrt{\mathsf{SNR}_n} \sum_{k=1}^{K} \sqrt{g_{nk}} h_{nk} x_k + \sqrt{\frac{\mathsf{SNR}_n}{\mathsf{SIR}_n}} z_n' + z_n \qquad n = 1, \ldots, N \tag{8}$$

where the noise terms $\{z_n\}$, the out-of-cluster interference terms $\{z_n'\}$, the signals $\{x_k\}$, and the fading channel coefficients $\{h_{nk}\}$ are all mutually independent random variables normalized to be unit-variance.

If both the noise and the out-of-cluster interference are Gaussian, then (8) becomes

$$y_n = \sqrt{\mathsf{SNR}_n} \sum_{k=1}^{K} \sqrt{g_{nk}} h_{nk} x_k + \sqrt{1 + \frac{\mathsf{SNR}_n}{\mathsf{SIR}_n}} z_n'' \qquad n = 1, \ldots, N \tag{9}$$

where z_n'' is the aggregate noise-plus-interference, also Gaussian and unit variance, i.e., $z_n'' \sim \mathcal{N}_{\mathbb{C}}(0,1)$. Since the Gaussian distribution correctly models thermal noise, and the out-of-cluster interference is made up of a large number of independent terms and thus its distribution tends to be approximately Gaussian too, we focus on (9) with $z_n'' \sim \mathcal{N}_{\mathbb{C}}(0,1)$. Nevertheless, all of the points made henceforth apply (qualitatively) in the wider generality of (8), with z_n and z_n' having different distributions. Eq. (9) can be further rewritten into the following.

Relationship 3 *The observation at receiver n is given by*

$$y_n = \sqrt{\frac{\mathsf{SNR}_n \, \mathsf{SIR}_n}{\mathsf{SNR}_n + \mathsf{SIR}_n}} \sum_{k=1}^{K} \sqrt{g_{nk}} h_{nk} x_k + z_n'' \tag{10}$$

$$= \sqrt{\mathsf{SINR}_n} \sum_{k=1}^{K} \sqrt{g_{nk}} h_{nk} x_k + z_n'' \qquad n = 1, \ldots, N \tag{11}$$

where SINR_n is the signal-to-interference-plus-noise ratio at receiver n, which equals the harmonic mean of SNR_n and SIR_n, i.e.,

$$\frac{1}{\mathsf{SINR}_n} = \frac{1}{\mathsf{SNR}_n} + \frac{1}{\mathsf{SIR}_n}. \tag{12}$$

For small P, then, $\text{SINR}_n \approx \text{SNR}_n$ whereas for $P \to \infty$, $\text{SINR}_n \to \text{SIR}_n$. The formulation on the basis of SNR_n and SIR_n is very general in that it captures not only the scaling with P, but also with other parameters such as cell size or noise variance.

B. Small-Scale Modeling

Referring back to Relationship 3, the small-scale fading is modeled as Rayleigh and thus the normalized fading coefficients satisfy $h_{nk} \sim \mathcal{N}_{\mathbb{C}}(0,1)$. We consider frequency-selective fading with coherence bandwidth B_{c} and thus the value of each h_{nk} varies from subband to subband in an IID (independent identically distributed) fashion. In terms of the temporal fading dynamics, both block- and continuous-fading are accommodated.

- With block fading, the channels within each subband hold constant for some time and then change to a different value also in an IID fashion. If we denote by T_{c} the channel coherence in time, then—irrespective of how the signaling is arranged along the time and frequency dimensions—the number of symbols over which the channel remains coherent is roughly $L = B_{\mathrm{c}}T_{\mathrm{c}}$.
- With continuous fading, the channels within each subband are discrete-time stationary and ergodic random processes with a Doppler spectrum $S_h(\cdot)$ that is bandlimited, i.e.,

$$\begin{cases} S_h(\nu) > 0 & |\nu| \leq f_{\mathrm{D}} \\ S_h(\nu) = 0 & |\nu| > f_{\mathrm{D}} \end{cases} \tag{13}$$

for some maximum Doppler frequency $f_{\mathrm{D}} \leq 1/2$. In cellular channels, $f_{\mathrm{D}} = v/(\lambda B_{\mathrm{c}})$, where v is the velocity and λ is the carrier wavelength.

Block fading is a coarse but effective approximation to continuous fading, and remarkable equivalences between the two have been uncovered [26], [27]. In particular, both models have been shown to be equivalent in terms of channel estimation MMSE (minimum mean-square error) for the case of a rectangular Doppler spectrum, $S_h(\nu) = \frac{1}{2f_{\mathrm{D}}}$ for $|\nu| \leq f_{\mathrm{D}}$. With such a spectrum, the MMSE is equivalent to that of a block-fading channel with $L = \frac{1}{2f_{\mathrm{D}}}$ (cf. Section III).

Example 1 *With the typical cellular values $\lambda = 0.15$ m and $B_{\mathrm{c}} = 370$ KHz, $f_{\mathrm{D}} \approx 2.5 \times 10^{-5}$ for pedestrian velocities and $f_{\mathrm{D}} \approx 5 \times 10^{-4}$ for vehicular velocities. This maps to $L \approx 20{,}000$ and $L \approx 1000$, respectively.*

C. Exemplary Cellular System

To close this section, let us introduce an exemplary system that shall be utilized throughout.
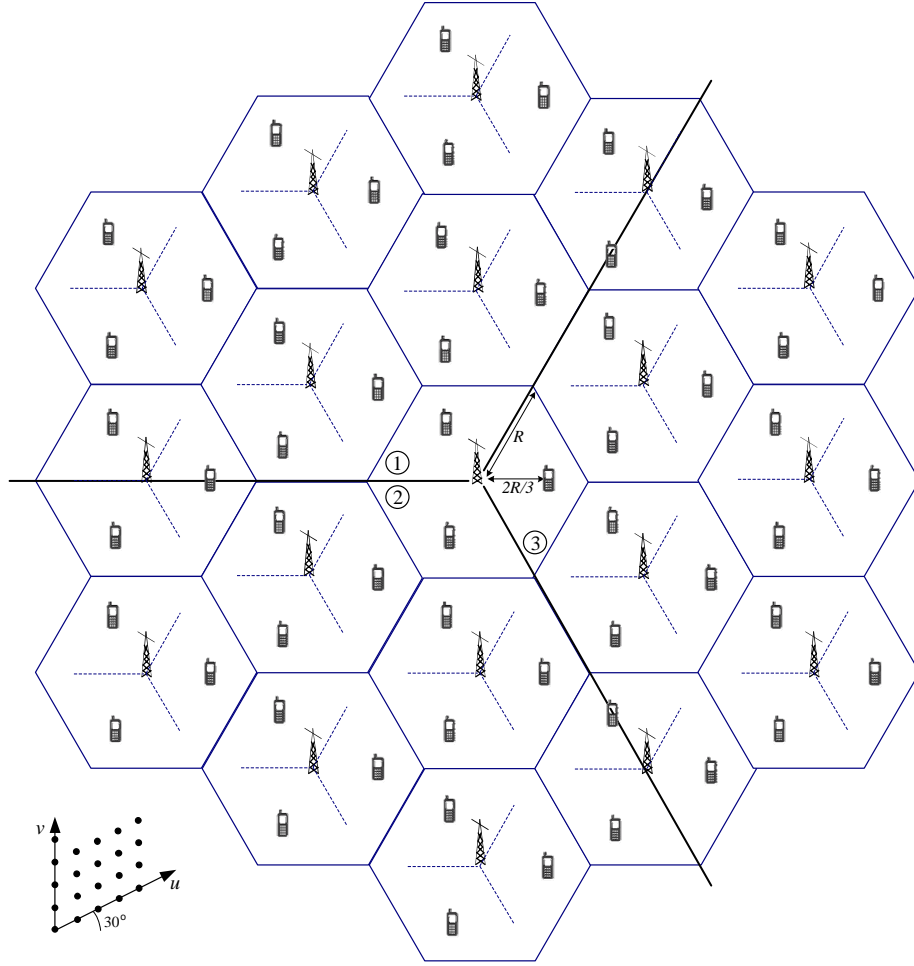
Fig. 3. Regular hexagonal universe with tri-sector hexagonal cells of size R. Each user is centered within its sector and at distance $2R/3$ from the BS.

Example 2 *Consider a cellular system with tri-sector hexagonal cells of size R. Each sector's antenna has a uniform gain over the $120°$ span of the sector and a $Q|_{\mathrm{dB}}$-lower uniform gain outside that span. Orthogonal signaling resources (time slots and frequency subcarriers) are allocated to the users within each sector. On any given resource, thus, there is a single user per sector and hence $K = N$. Each user is centered in azimuth within its sector and at distance $2R/3$ from the BS. Depicted in Fig. 3 are an arbitrary reference cell and the two tiers around it, yet the system has infinitely many cells. The signals experience distance-dependent decay with an exponent γ and also Rayleigh fading.*

Example 2 is representative of a cellular system, while having the virtue of being isotropic and of having regular user locations. Eq. (9) can be embodied for any desired values of $K = N$, i.e., for clusters of arbitrary size, and the out-of-cluster interference can be easily summed for $\tilde{K}, \tilde{N} \to \infty$ thanks to the regular user locations.

III. PILOT-BASED CHANNEL ESTIMATION

In this section we establish that the spectral efficiency inevitably saturates for large P, under the restriction that the receivers estimate the channel coefficients on the basis of pilot symbols. We shall remove this restriction in the sequel. The rationale for studying systems featuring explicit channel estimation followed by coherent detection of payload data [28] is that virtually every existing wireless system abides by this procedure. References on cooperation with pilot-based channel estimation include [21], [29]–[33].

The procedure entails the transmission of pilot symbols regularly in time (and in frequency when multiples of B_c are spanned). We note that (*i*) pilot symbols are overhead, and (*ii*) the periodicity of pilot transmission is determined by the channel coherence. Only a finite number of pilot symbols can be transmitted within a given coherence interval containing L symbols. At least one pilot symbol, and most likely multiple ones, must be devoted to enabling the estimation of each channel coefficient at the corresponding receiver for each coherence interval. If the channels were matrix-valued, as in the MIMO case, separate pilot symbols would be needed for every entry therein. The number of channel coefficients that can be estimated is therefore limited by L and must be substantially lower than L for the overhead not to be overwhelming. Explicit channel estimation can thus occur only within clusters of limited dimension, inevitably with out-of-cluster interference from all the transmitters beyond. Excessively large clusters incur excessive overhead and/or poor channel estimates that ultimately nullify the benefits of cooperation.

To make this intuition specific, denote by α the share of the symbols reserved for pilots; the rest, $(1-\alpha)$, is for payload data. The pilot transmissions should be orthogonally multiplexed from each of the transmitters [34] and thus the estimation of each of the K channel coefficients at each receiver relies on a share α/K of the symbols. When only the kth transmitter is actively sending pilots, the SINR at receiver n is $g_{nk}\mathsf{SINR}_n$. With block fading, the MMSE on the estimation of h_{nk} is then [28]

$$\mathsf{MMSE}_{nk} = \frac{1}{1 + g_{nk}\mathsf{SINR}_n L\alpha/K} \tag{14}$$

whereas, with continuous fading [26]

$$\mathsf{MMSE}_{nk} = 1 - \int_{-f_{\mathrm{D}}}^{f_{\mathrm{D}}} \frac{g_{nk}\mathsf{SINR}_n S_h^2(\nu)}{K/\alpha + g_{nk}\mathsf{SINR}_n S_h(\nu)} d\nu. \tag{15}$$

If the Doppler spectrum $S_h(\cdot)$ is rectangular, (15) becomes

$$\mathsf{MMSE}_{nk} = \frac{1}{1 + g_{nk}\mathsf{SINR}_n \frac{\alpha/K}{2f_{\mathrm{D}}}} \tag{16}$$

which coincides with (14) for $f_{\mathrm{D}} = \frac{1}{2L}$. The MMSE expression in (16) thus allows embracing both the block- and continuous fading models in a single framework.

Denoting the estimate by $\hat{h}_{nk}$ and the estimation error by $\tilde{h}_{nk}$, it follows that $h_{nk} = \hat{h}_{nk} + \tilde{h}_{nk}$ with $\hat{h}_{nk}$ and $\tilde{h}_{nk}$ uncorrelated and with $\mathbb{E}[|\tilde{h}_{nk}|^2] = \mathsf{MMSE}_{nk}$. Hence, from Relationship 3,

$$y_n = \sqrt{\mathsf{SINR}_n} \sum_{k=1}^{K} \sqrt{g_{nk}} \hat{h}_{nk} x_k + \sqrt{\mathsf{SINR}_n} \sum_{k=1}^{K} \sqrt{g_{nk}} \tilde{h}_{nk} x_k + z_n'' \qquad n = 1, \ldots, N. \qquad (17)$$

Typically the receivers apply minimum-distance decoding utilizing the channel estimates as if they were correct, in which case the terms in the second summation in (17) play the role of additional Gaussian noise [35]. With that, the effective SINR at receiver n upon payload data detection is

$$\mathsf{SINR}_n^{\mathsf{eff}} = \frac{\mathsf{SINR}_n \sum_{k=1}^{K} g_{nk}(1 - \mathsf{MMSE}_{nk})}{1 + \mathsf{SINR}_n \sum_{k=1}^{K} g_{nk}\mathsf{MMSE}_{nk}} \qquad (18)$$

and the average spectral efficiency (bits/s/Hz/user) that can be attained reliably is

$$C = (1 - \alpha)\frac{1}{N} f\left(\mathsf{SINR}_1^{\mathsf{eff}}, \cdots, \mathsf{SINR}_N^{\mathsf{eff}}\right) \qquad (19)$$

where the function $f(\cdot)$ depends on the type of cooperation among the $K = N$ users. The spectral efficiency, which vanishes for both $\alpha \to 0$ and $\alpha \to 1$, is maximized by a proper choice of $0 < \alpha < 1$.

As K grows, α needs to grow at least linearly with it to maintain the effective SINRs. Since enlarging the cooperation clusters entails involving progressively weaker signals, α might have to grow superlinearly with K. Ultimately, C is bound to peak at some cluster size and diminish thereafter.

Example 3 *Consider the uplink of the system in Example 2 with clusters of arbitrary size $K = N$ where the BSs fully cooperate via Network MIMO, i.e., they jointly decode the K signals received at the N sectors. The distance-decay exponent is $\gamma = 3.8$ whereas $Q|_{\mathrm{dB}} = 20$ dB. Defining an $N \times K$ matrix $\mathbf{S}$ whose entries are independent and such that the (n, k)th entry is $\mathbf{S}_{nk} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_{nk}^2)$ where*

$$\sigma_{nk}^2 = \frac{g_{nk}\mathsf{SINR}_n(1 - \mathsf{MMSE}_{nk})}{1 + \mathsf{SINR}_n \sum_{\kappa=1}^{K} g_{n\kappa}\mathsf{MMSE}_{n\kappa}}, \qquad (20)$$

the average spectral efficiency (bits/s/Hz/user) is

$$C = (1 - \alpha)\frac{1}{N}\mathbb{E}[\log_2 \det(\mathbf{I} + \mathbf{S}\mathbf{S}^{\dagger})]. \qquad (21)$$

Note that, because of isotropy, $\mathsf{SNR}_n = \mathsf{SNR}$ for $n = 1, \ldots, N$. The fading is either block ($L = 20,000$) or continuous with a rectangular Doppler spectrum ($f_{\mathrm{D}} = 2.5 \times 10^{-5}$), which are equivalent in terms of channel estimation and correspond to pedestrian velocities. Shown in Fig. 4 is the spectral efficiency for several cluster sizes. The baseline $K = N = 1$ corresponds to single-user decoding, i.e., no cooperation. The case $K = N = 3$ corresponds to a cluster of 3 facing sectors as in Fig. 3. Finally, the case $K = N = 21$ corresponds to a cluster of 7 cells: one central cell plus the first tier around it. The pilot overhead α is equal for all users but otherwise it is optimized to maximize C, i.e., it is optimized
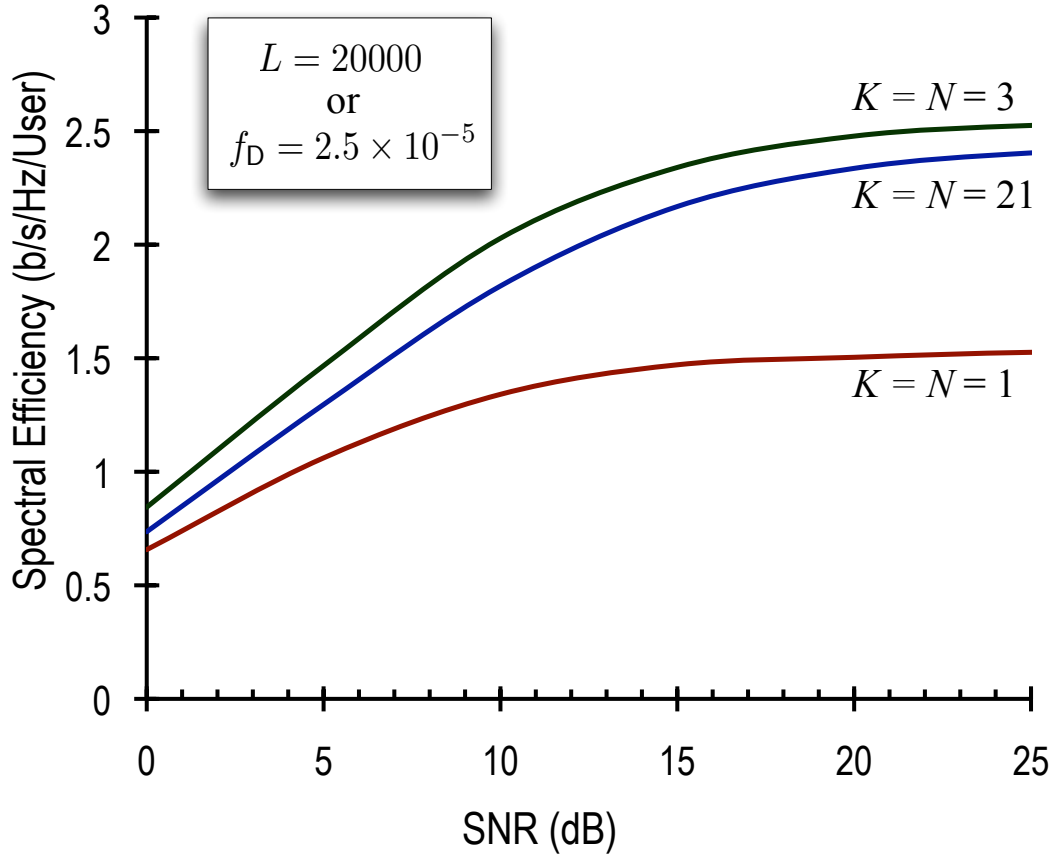
Fig. 4.   Spectral efficiency as function of SNR for varying cluster sizes with either block fading ($L = 20,000$) or continuous fading ($f_{\rm D} = 2.5 \times 10^{-5}$). Uplink with full cooperation.

for every cluster size and power level. For details on how the out-of-cluster interference is computed, see Appendix C.

The performance in Example 3 improves when the cluster size goes from 1 to 3, but then degrades when the cluster size increases to 21 and continues to degrade for even larger sizes. To gauge the impact of out-of-cluster interference, we re-plot the spectral efficiency corresponding to $K = N = 3$ in Fig. 4 next to that obtained in the same exact conditions only with all the transmitters outside the cluster turned off. This comparison, presented in Fig. 5, evidences that without out-of-cluster interference we recover the traditional behavior (cf. Fig. 1). Out-of-cluster interference, however, drastically modifies that behavior (cf. Fig. 2). Note that modeling out-of-cluster interference as additional noise of fixed variance cannot fix the representation in Relationship 1, as it merely would shift the spectral efficiency by some fixed amount. Only the representation in Relationships 2–3 can properly reproduce the correct behavior.

Example 4 *For Example 3 with $K = N = 3$, the SIR that should be inserted into Relationship 3 can*
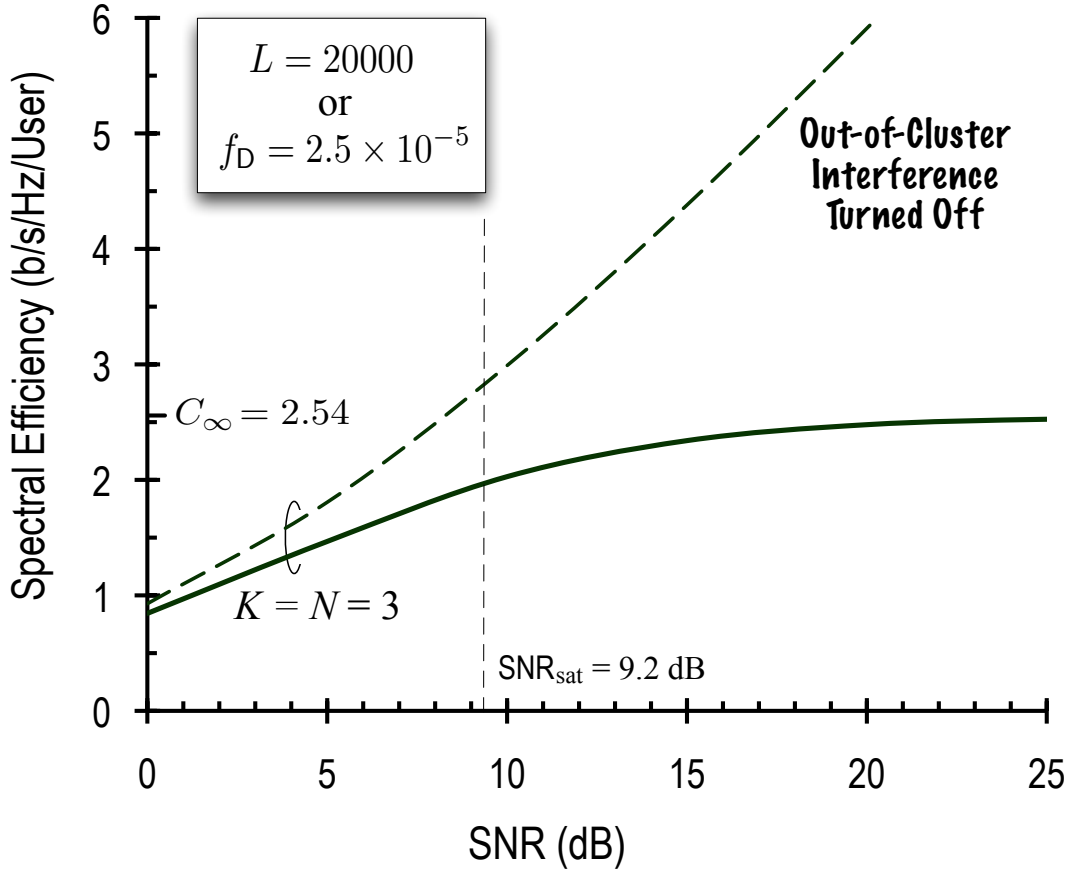
Fig. 5. Spectral efficiency as function of SNR for $K = N = 3$ with either block fading ($L = 20,000$) or continuous fading ($f_D = 2.5 \times 10^{-5}$). Uplink with full cooperation. In solid, with out-of-cluster interference included. In dashed, with out-of-cluster interference turned off.

be found to be

$$\mathsf{SIR}_n = \frac{\sum_{k=1}^{3} G_{nk}}{\sum_{k=4}^{\infty} G_{nk}} \tag{22}$$

$$= 9.2 \text{ dB} \qquad n = 1, 2, 3 \tag{23}$$

which, indeed, corresponds with the inflection point observed in Fig. 5. That point, which corresponds to P_{sat} and thus is denoted by $\mathsf{SNR}_{\mathsf{sat}}$, delineates the transition between the DoF and the saturation regimes. In turn, the limiting spectral efficiency can be found to be $C_\infty = 2.54$ bits/s/Hz/user.

For user locations different from the regular ones in Example 3, the values of $\{\mathsf{SIR}_n\}$ vary and thus the transition between the DoF and the saturation regimes takes place at different points, but qualitatively speaking the behavior is unaltered.

Example 5 *Consider a variation of Example 3, with the out-of-cluster users still centered within their own sectors but with the location of the in-cluster users randomized. Shown in Fig. 6 is the cumulative*
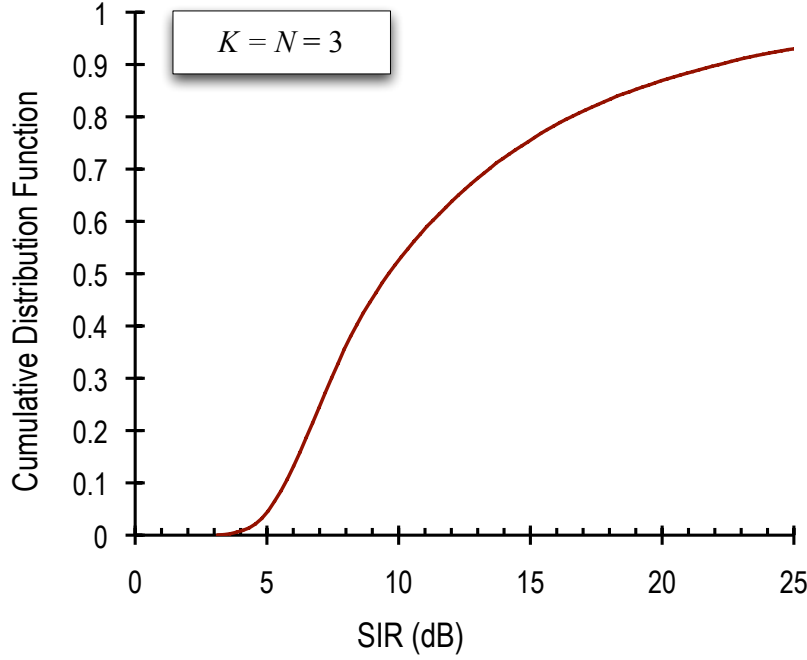
Fig. 6.   Cumulative distribution of SIR (dB) over all in-cluster locations for $K = N = 3$. Uplink with $K = N = 3$ and full cooperation.

distribution of SIR at each of the three in-cluster receivers. In almost 90% of locations, the SIR is below 20 dB. The median SIR is at 9.6 dB, confirming that the value obtained in Example 4 is indeed quite representative of the average conditions.

The figures derived in the preceding example indicate that, in a vast majority of cases, the spectral efficiency saturates at SNR levels of operational interest. The final steps to fully generalize these figures would be to randomize the location of the out-of-cluster users and to incorporate shadow fading on all the links. Exactly summing the out-of-cluster interference becomes challenging in that broad generality, but the computation is otherwise conceptually identical and the resulting SIR distribution is not expected to depart much from the one in Fig. 6.

Qualitatively similar observations to the ones made in this section for the uplink can be made for the downlink, which could be analyzed by extending the pilot-assisted scheme in [36] to the multicell realm.

IV. NONCOHERENT DETECTION

In this section we consider noncoherent detection without explicit channel estimation. From an information-theoretic perspective, noncoherent detection subsumes as a special case the procedure of transmitting pilot symbols, estimating the channel coefficients, and detecting the data coherently. Pilot

symbols are a specific form of redundancy and explicit channel estimation followed by coherent data detection is a specific detection strategy. It could be argued that the need for clusters is not fundamental but simply a byproduct of having limited opportunities (due to the finite coherence) for pilot transmission. If explicit channel estimation could be transcended by means of noncoherent detection, then perhaps the entire system could cooperate as one. Following this argument, there would be no out-of-cluster interference and Relationship 1 would remain a valid representation. In this section we refute this idea and show that, in large networks, the saturation of the spectral efficiency is still unavoidable and Relationship 3 remains the correct representation for a cellular system or any cluster therein.

In the absence of CSI (channel-state information) at the receivers, the capacity-achieving signals and the capacity itself are generally unknown—even for a single-user channel. The spectral efficiency achievable with complex Gaussian signals is also unknown, save for MIMO channels with IID entries [37]. Given the difficulty in computing the exact spectral efficiency achievable without CSI at the receivers, we instead show that its limiting value for $P \to \infty$, C_∞, is bounded by above by a quantity that does not depend on P; hence, C cannot grow without bound with P.

With clusters of limited dimension, the presence of out-of-cluster interference is sure to bring about a finite C_∞ with noncoherent detection just as it did with explicit channel estimation. The question that lingers is whether this changes for $K = \tilde{K}$ and $N = \tilde{N}$, i.e., when the entire system cooperates as one.

Consider the uplink. With block fading, the transmit-receive relationship can be vectorized for the entire system and all the symbols in a fading block as

$$\mathbf{Y} = \mathrm{diag}\left\{\sqrt{\mathsf{SNR}_1}, \ldots, \sqrt{\mathsf{SNR}_N}\right\} \mathbf{H}\mathbf{X} + \mathbf{Z} \tag{24}$$

where $\mathbf{Y}$ and $\mathbf{Z}$ are $N \times L$, $\mathbf{H}$ is $N \times K$ and $\mathbf{X}$ is $K \times L$. The entries of $\mathbf{Z}$ are IID with $\mathbf{Z}_{n\ell} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ whereas the entries of $\mathbf{H}$ are independent with $\mathbf{H}_{nk} \sim \mathcal{N}_{\mathbb{C}}(0, g_{nk})$; notice that the normalized channel power gains $\{g_{nk}\}$ are directly incorporated as the variances of the entries of $\mathbf{H}$. The entries of $\mathbf{X}$ are unit variance and, since the kth row of $\mathbf{X}$ contains the signal sequence transmitted by user k over the L symbols of a fading block, the rows of $\mathbf{X}$ are independent. Since the entire system is represented in (24), there is no out-of-cluster interference.

Proposition 1 *Consider the uplink of a cellular system subject to block-fading with $K > L$. Define, for $n = 1, \ldots, N$, respective diagonal matrices $\mathbf{G}_n = \mathrm{diag}\{g_{n1}, \ldots, g_{nK}\}$. Each $\mathbf{G}_n$ contains along its diagonal the normalized power gains between the K transmitters and the nth receiver and thus $\mathrm{Tr}\{\mathbf{G}_n\} = 1$. If $\mathbf{X}$ is full rank, the average spectral efficiency that can be achieved reliably for $P \to \infty$ satisfies $C_\infty \leq C_\infty^{\mathsf{UB}}$ with*

$$C_\infty^{\mathsf{UB}} = -\frac{1}{K}\sum_{n=1}^{N}\frac{1}{L}\,\mathbb{E}\left[\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right)\right]. \tag{25}$$

Proof: *See Appendix A.*

In light of (5), C_∞^{UB} remains bounded for $K, N \to \infty$. In sufficiently large systems, therefore, there is no hope for C_∞ to grow without bound with P.

Note also that the rank constraint on $\mathbf{X}$ is very mild, accommodating every signaling strategy utilized in wireless communication. In particular, if the signal sequence transmitted by each user is IID complex Gaussian, then C_∞^{UB} in (25) can be expressed using [38, Prop. 4] in a form that is closed although not particularly convenient to work with when K and L are large.

A more compact analytical handle on C_∞^{UB} can be obtained by resorting to large-dimensional results in random matrix theory, with the added advantage that the expressions then hold for IID signals regardless of the distribution from which the symbols are drawn (e.g., PSK or QAM in addition to complex Gaussian).

Proposition 2 *Consider the uplink of a cellular system subject to block-fading with $K > L$. Then,*

$$C_\infty^{\text{UB}} \approx \frac{N}{K} \log_2 e + \frac{1}{K} \sum_{n=1}^{N} \left[\log_2 \frac{a_n}{L} - \frac{1}{L} \sum_{k=1}^{K} \log_2(1 + a_n g_{nk}) \right] \tag{26}$$

with each a_n the nonnegative solution to

$$\sum_{k=1}^{K} \frac{g_{nk}}{g_{nk} + 1/a_n} = L \tag{27}$$

and with (26) becoming exact as $K, L \to \infty$.

Proof: *See Appendix B.*

In certain special cases such as Example 2, Propositions 1 and 2 simplify significantly.

Corollary 1 *If the system is isotropic in the sense that the set of channel power gains from the transmitters looks the same from the vantage of each receiver, i.e., every set $\{g_{nk}\}$ for $n = 1, \ldots, N$ can be reordered into a common set $\{g_k\}$, then we can define a unique matrix $\mathbf{G}$ containing such set on its diagonal and rewrite (25) as*

$$C_\infty^{\text{UB}} = -\frac{N}{K} \frac{1}{L} \mathbb{E} \left[\log_2 \det \left(\mathbf{X}^\dagger \mathbf{G} \mathbf{X} \right) \right]. \tag{28}$$

In turn, Proposition 2 then specializes to

$$C_\infty^{\text{UB}} \approx \frac{N}{K} \left[\log_2 e + \log_2 \frac{a}{L} - \frac{1}{L} \sum_{k=1}^{K} \log_2(1 + a g_k) \right] \tag{29}$$

with the unique a being the nonnegative solution to

$$\sum_{k=1}^{K} \frac{g_k}{g_k + 1/a} = L. \tag{30}$$

In the context of continuous fading channels, the expressions in both Proposition 2 and Corollary 1 can be interpreted using the equivalence $L = \frac{1}{2f_D}$ put forth earlier in the paper.

Before applying Corollary 1 to the infinitely large system in Example 2, we verify the accuracy of the asymptotic approximation in (29) for finite $K = N$ and for L such that its non-asymptotic counterpart in (28) can be computed numerically.

Example 6 *Consider the uplink of a square fragment of the system in Example 2 having 20×20 cells ($K = 1200$) and let $L = 100$. All the sectors in the system cooperate fully via Network MIMO. The distance-decay exponent is $\gamma = 3.8$ whereas $Q|_{\mathrm{dB}} = 20$ dB. Monte-Carlo evaluation of (28) gives $C_\infty^{\mathrm{UB}} = 5.183$ b/s/Hz/user whereas (29) gives $C_\infty^{\mathrm{UB}} = 5.181$ b/s/Hz/user.*

Having verified its accuracy, let us apply Corollary 1 to the full system in Example 2 and for values of L in the range of practical interest.

Example 7 *Reconsider Example 6 for $K = N \to \infty$. For $L = 20,000$, $C_\infty^{\mathrm{UB}} = 11.86$ bits/s/Hz/user whereas, for $L = 1000$, $C_\infty^{\mathrm{UB}} = 7.98$ bits/s/Hz/user. For details, see Appendix C.*

The spectral efficiencies actually achievable for $P \to \infty$ may be substantially lower than the values in Example 7, not only because those correspond to upper bounds but also because the values of L used for the computations, which are correct over short distances, become optimistic in the context of cooperation across an entire system. The long propagation delays that arise when distant units cooperate is sure to lead to a longer delay spread and thus a smaller B_c, with the consequent reduction in L.

The exact saturation point notwithstanding, the spectral efficiency does saturate and thus Relationship 1 cannot represent an entire system operating as one. In contrast, Relationship 3 does adequately model the saturation that occurs for large P. A question that can be posed at this point is the following: if one wants to assume perfect CSI at the receivers, which values for $\{\mathrm{SIR}_n\}$ should be inserted into Relationship 3 to reproduce the results in Example 7? We can gauge this from the values of C_∞^{UB} obtained therein. Because of the isotropy in Example 7, every user operates at the same spectral efficiency and thus $\mathrm{SIR}_n = \mathrm{SIR}$ for $n = 1, \ldots, N$. With perfect CSI at the receivers, complex Gaussian signals are capacity-achieving and the uplink capacity of Relationship 3 with full cooperation equals, for $P \to \infty$,

$$C_\infty = \frac{1}{K} \log_2 \det \left(\mathbf{I} + \mathrm{SIR}\, \mathbf{H}\mathbf{H}^\dagger \right). \tag{31}$$

The structure of $\mathbf{H}$ satisfies the conditions of [39, Thm. 5] and thus, for $K = N \to \infty$,

$$C_\infty \overset{\text{a.s.}}{\to} 2\log_2 \left(\frac{1 + \sqrt{1 + 4\,\mathrm{SIR}}}{2} \right) - \frac{\log_2(e)}{4\,\mathrm{SIR}} \left(\sqrt{1 + 4\,\mathrm{SIR}} - 1 \right)^2. \tag{32}$$

Solving for the sɪʀ that equates (32) with the values for C_∞^{UB} in Example 7 we obtain

$$\text{sɪʀ} = 39.96 \, \text{dB}, \quad L = 20,000 \, \text{(pedestrian)}$$
$$\text{sɪʀ} = 28.02 \, \text{dB}, \quad L = 1000 \, \text{(vehicular)}$$
(33)

which, deriving from spectral efficiency upper bounds, are themselves upper bounds on the actual sɪʀ. Although, with shadow fading and randomized user locations, these values are likely to vary significantly, they are indicative: even if an entire system were to be operated as one with full cooperation, interference brings about a fundamental performance ceiling that corresponds to values of sɪʀ within the range of interest in high-power analysis. This ceiling, furthermore, depends exclusively on the coherence L, which relates to the degree of mobility, and the geometry profile $\{g_{nk}\}$, which quantifies the degree of signal connectivity among users.

Besides computational convenience, an additional benefit of the asymptotic approximations in Proposition 2 and Corollary 1 is that they cast light on how C_∞^{UB} depends on the geometry profile.

- If $\{g_{nk}\}$ is highly skewed for a given n, then most of the power received by BS n corresponds to a few nearby users. Intuition then says that, given their relative strength, the fading of these users' channels and the overlaying signals could be determined and most of the received power should be rendered useful. The asymptotic results confirm this intuition: with fewer than L nonnegligible terms, C_∞^{UB} can be arbitrarily large and thus a sustained increase of the spectral efficiency with P is feasible.

- Alternatively, if $\{g_{nk}\}$ for a given n contains a myriad minute terms, rather than a few strong ones, each of these terms is simply too weak relative to the aggregate rest. Intuitively, this should give rise to a bulk of residual interference that is fundamentally undecodable, and which is substantial in the aggregate even if each of the terms is by itself small. Again, the asymptotic results confirm this intuition: if $g_{nk} = 1/K$ for $k = 1, \ldots, K$, then, as K grows without bound for fixed L, the interference becomes overwhelming and C_∞^{UB} vanishes.

In actual systems, there will be a few strong signals in a sea of minute ones. The result is a finite value for C_∞^{UB} that depends on the skewness of the geometry profile. Note further that, since the $\{g_{nk}\}$ are normalized, they are scale independent. Cell size is therefore immaterial in terms of the geometry profile.

Note also that the schedulers that determine which user(s) in each cell are allocated to a given signaling resource play a significant role in establishing the geometry profile. Subject to latency and quality-of-service constraints, the schedulers can therefore shape the geometry profile. Dynamic definition of the cooperation clusters is also likely to be beneficial [40], [41]. Other aspects that may affect it include (possibly fractional) frequency reuse, power control, antenna patterns, and antenna downtilting [42].

For the downlink, the roles of transmission and reception and reversed relative to the uplink. The corresponding noncoherent performance can be upper-bounded by allowing all the receivers to cooperate, in which case the uplink derivations carry over and a performance ceiling is readily observed. Tighter upper bounds might be obtained by removing the premise of receiver cooperation while considering downlink transmission strategies other than IID signaling.

V. CONCLUDING DISCUSSION

As argued up to this point, Relationship 3 with the appropriate values for $\{\mathsf{SIR}_n\}$ is the correct representation of a cellular system, or any fragment thereof. The traditional high-power regime for user n, characterized by $\mathsf{SNR}_n \gg 1$, splits into two regimes (cf. Fig. 2):

1) The DoF regime where $\mathsf{SNR}_n \ll \mathsf{SIR}_n$. In this regime, the out-of-cluster interference is negligible relative to the noise and the spectral efficiency grows approximately linearly with $\log(P)$ according to the number of DoF computed without out-of-cluster interference. The notion of DoF remains a valid approximation.

2) The saturation regime where SNR_n is comparable to or greater than SIR_n. In this regime, the spectral efficiency chokes as it approaches C_∞. The notion of DoF becomes meaningless; more precisely, the number of DoF is revealed to be zero.

The transition between these regimes takes place, for user n, at $\mathsf{SNR}_{\mathsf{sat},n} \approx \mathsf{SIR}_n$. If the value of SIR_n is not explicitly known but C_∞ is known, the transition can alternatively be ascertained by assessing where an interference-free high-power expansion of the spectral efficiency intersects with C_∞. It is pointless to operate much above $\mathsf{SNR}_{\mathsf{sat}}$.

A. The Benefits of Cooperation

We emphasize that the points made in this paper do not nullify the benefits of cooperation but, rather, they show that cooperation has fundamental limitations that cannot be overcome through faster backhaul, more sophisticated signal processing, or any other technological advance. Under Relationship 3, cooperation can still yield a markedly higher C_∞ than if all interference was simply ignored. Similarly, cooperation can provide an increased slope within the DoF regime[1], allowing C_∞ to be approached at lower power levels. This is demonstrated in the following example.

Example 8 *Consider Relationship 3 with $K = N = 3$ and let every BS and user have two antennas. Further let $\mathsf{SNR}_n = \mathsf{SNR}$, $n = 1, 2, 3$. Shown in Fig. 7 are the spectral efficiencies achieved by distributed*

---

[1]Recall that, although the number of DoF is strictly speaking zero, in this regime the slope is well approximated by the number of DoF with the out-of-cluster interference neglected
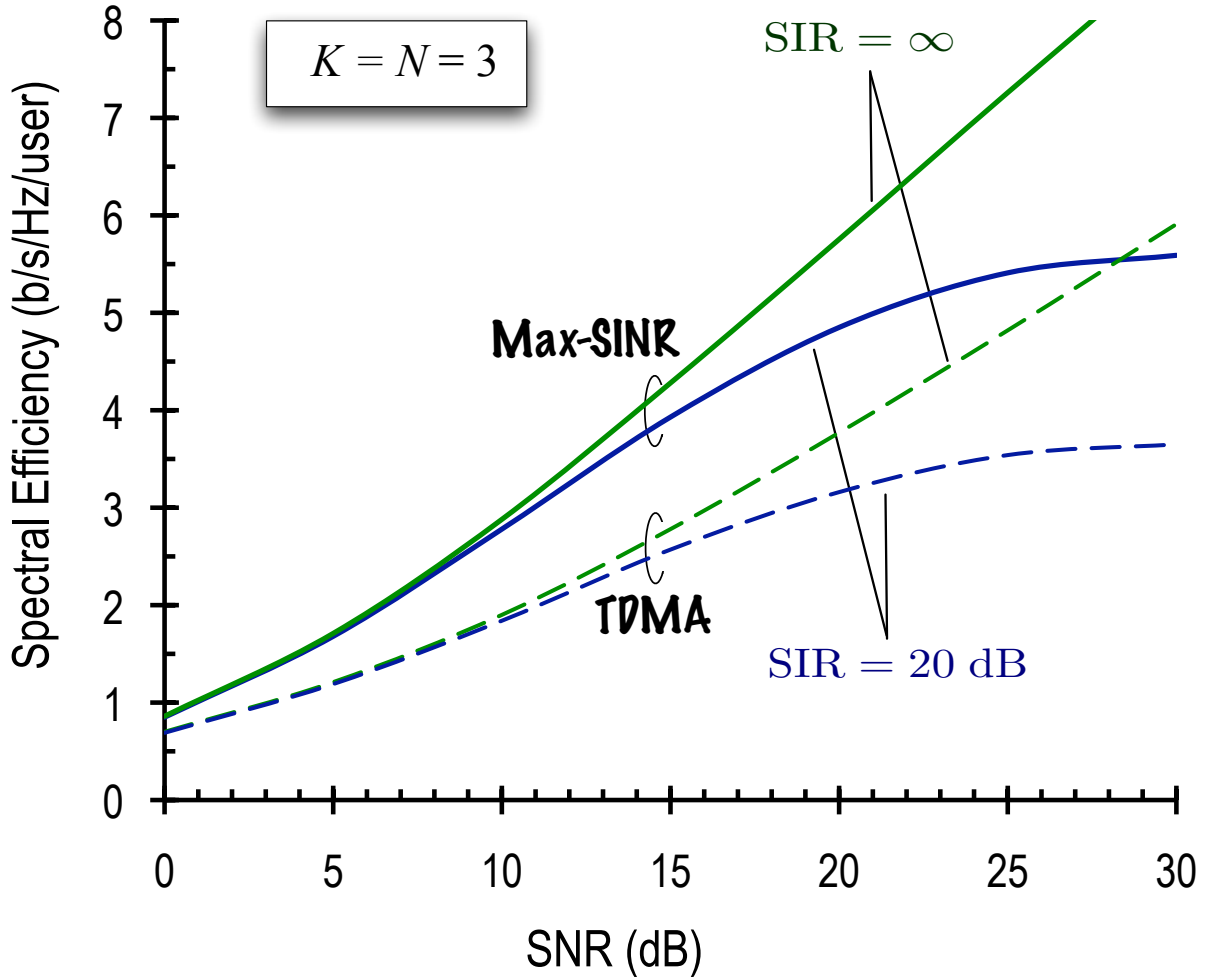
Fig. 7. Max-SINR v. TDMA using Relationship 3 with $K = N = 3$ and with two antennas per transmitter and per receiver, for SIR $= \infty$ and SIR $= 20$ dB.

Max-SINR [15], [16] and by round-robin TDMA for both SIR $= \infty$ *(no out-of-cluster interference) and for* SIR $= 20$ *dB. Note that, for* SIR $= \infty$, *Relationship 3 reverts to Relationship 1 and the theoretical number of DoF per user (1 with Max-SINR and $2/3$ with TDMA) are approached for* SNR $\to \infty$. *For* SIR $= 20$ *dB, the inflection that delineates the DoF and saturation regimes occurs around* SNR $= 20$ *dB as expected.*

Apropos Example 8 with SIR $= 20$ dB, we observe that the value of C_∞ is over 50% higher with Max-SINR relative to TDMA. Further, in the DoF regime—from about 5 to 20 dB—a substantial difference between Max-SINR and TDMA builds up. Altogether then, cooperation continues to provide a substantial advantage but the number of DoF is only a partial measure thereof.

B. Future Directions

This paper has shown that saturation of the spectral efficiency at sufficiently high powers is unavoidable in large systems. A key insight is that interference is made up of a multitude of terms, the majority of which are too minute to be tracked. Beyond a few strong terms, the structure of the interference is too intricate to discern yet the sum of all these minute terms is significant. Essentially, the receivers are near-sighted. They can only focus on a few strong nearby transmitters, and everything else in the distance looks fundamentally blurry. The lack of focus depends on the extent of mobility in the system.

Further work is needed to expand the observations in this paper. For example, the analysis focused on the uplink; the extension to the downlink is interesting and nontrivial. Only single-antenna transmitters and receivers were considered. It would be good to extend the results to MIMO, including large-dimensional MIMO systems that are limited by pilot contamination [43]. In particular, it would be of interest to understand the tradeoffs between numbers of antennas at a single base station and coordination of many base stations to understand if there is any fundamental preference for centralized or distributed antenna architectures. If the system is small or the clusters are highly isolated (e.g., through geographical separation, penetration losses, or millimeter-wave propagation losses) then the $\mathsf{SNR_{sat}}$ can be so large as to render the saturation anecdotal. It would be useful to understand and classify systems based on the proximity of $\mathsf{SNR_{sat}}$ to practical operating SNRs say less than 30dB. The results in this paper neglected propagation delay among distant units. Characterizing the impact of delay (as a function of bandwidth) on $\mathsf{SNR_{sat}}$ and C_∞ would likely further reduce both of these values, making them more operationally relevant. Finally, it would be good to know what an optimal cooperation cluster size is, in both the coherent and noncoherent realms.

APPENDIX

A. Proof of Proposition 1

Using the chain rule, the mutual information between $\mathbf{X}$ and $\mathbf{Y}$ (in bits per coherence block) can be expressed as

$$I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{HX}; \mathbf{Y}) - I(\mathbf{Y}; \mathbf{H}|\mathbf{X}). \tag{34}$$

Invoking (5) and the zero-mean unit-variance nature of the entries of $\mathbf{X}$, it can be verified that the entries of $\mathbf{HX}$ are also zero-mean and unit-variance. The term $I(\mathbf{HX}; \mathbf{Y})$ is then upper-bounded by the value it would take if those entries were IID $\mathcal{N}_{\mathbb{C}}(0, 1)$. Thus,

$$I(\mathbf{X}; \mathbf{Y}) \leq \sum_{n=1}^{N} L \log_2(1 + \mathsf{SNR}_n) - I(\mathbf{Y}; \mathbf{H}|\mathbf{X}). \tag{35}$$

Turning now our attention to $I(\mathbf{Y}; \mathbf{H}|\mathbf{X})$, and denoting differential entropy operator by $\mathfrak{h}(\cdot)$,

$$I(\mathbf{Y}; \mathbf{H}|\mathbf{X}) = \mathfrak{h}(\mathbf{Y}|\mathbf{X}) - \mathfrak{h}(\mathbf{Y}|\mathbf{H}, \mathbf{X}). \tag{36}$$

Conditioned on $\mathbf{X}$, the rows of $\mathbf{Y}$ are independent (both conditionally and unconditionally on $\mathbf{H}$) and thus both differential entropy terms in (36) can be computed row-wise and simply added. It follows that

$$I(\mathbf{Y}; \mathbf{H}|\mathbf{X}) = \sum_{n=1}^{N} I(\mathbf{y}_n; \mathbf{h}_n|\mathbf{X}) \tag{37}$$

where $\mathbf{y}_n = \sqrt{\mathsf{SNR}_n}\, \mathbf{h}_n \mathbf{X} + \mathbf{z}_n$ with $\mathbf{h}_n$ and $\mathbf{z}_n$ the nth rows of $\mathbf{H}$ and $\mathbf{Z}$, respectively. Since $\mathbf{h}_n$ and $\mathbf{z}_n$ are complex Gaussian vectors,

$$I(\mathbf{Y}; \mathbf{H}|\mathbf{X}) = \sum_{n=1}^{N} \mathbb{E}\left[\log_2 \det\left(\mathbf{I} + \mathsf{SNR}_n \mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right)\right] \tag{38}$$

where, recall, $\mathbf{G}_n = \mathrm{diag}\{g_{n1}, \ldots, g_{nK}\}$.

Since $I(\mathbf{X}; \mathbf{Y})$ increases monotonically with P, we concentrate on upper-bounding it for $P \to \infty$. Suppose first that we had $K < L$. Then, for large SNR_n, $n = 1, \ldots, N$, we would have

$$I(\mathbf{Y}; \mathbf{H}|\mathbf{X}) = \sum_{n=1}^{N} \left(K \log_2 \mathsf{SNR}_n + \mathbb{E}\left[\log_2 \det\left(\mathbf{X}\mathbf{X}^\dagger \mathbf{G}_n\right)\right]\right) + o(1) \tag{39}$$

and, from (35),

$$I(\mathbf{X}; \mathbf{Y}) \leq \sum_{n=1}^{N} \left((L - K) \log_2 \mathsf{SNR}_n - \mathbb{E}\left[\log_2 \det\left(\mathbf{X}\mathbf{X}^\dagger \mathbf{G}_n\right)\right]\right) + o(1) \tag{40}$$

whose right-hand side grows unboundedly with P indicating that there is hope for $I(\mathbf{X}; \mathbf{Y})$ to grow unboundedly with P.

Conversely, when $K \geq L$, and given the full-rank condition of $\mathbf{X}$,

$$I(\mathbf{Y}; \mathbf{H}|\mathbf{X}) = \sum_{n=1}^{N} \left(L \log_2 \mathsf{SNR}_n + \mathbb{E}\left[\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right)\right]\right) + o(1) \tag{41}$$

and thus (35) becomes, for large P,

$$I(\mathbf{X}; \mathbf{Y}) \leq -\sum_{n=1}^{N} \mathbb{E}\left[\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right)\right] + o(1) \tag{42}$$

from which the claimed upper bound in (25) follows.

B. Proof of Proposition 2

If each user transmits an IID sequence, then the entries of $\mathbf{X}$ are IID and the asymptotic analysis in [39], [44] can be applied to Proposition 1. Couched in the notation of this paper, we have that [39, Section VI-A]

$$-\frac{1}{L}\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right) \approx \log_2 e - \log_2\left(\sum_{k=1}^{K}\frac{g_{nk}}{1+a_{nk}}\right) - \frac{1}{L}\sum_{k=1}^{K}\log_2(1+a_{nk}) \tag{43}$$

where the approximation becomes exact as $K, L \to \infty$, with $K \geq L$, and where a_{nk} is the nonnegative solution to

$$a_{nk} = \frac{L\, g_{nk}}{\sum_{j=1}^{K}\frac{g_{nj}}{1+a_{nj}}}. \tag{44}$$

Defining

$$a_n = \frac{L}{\sum_{j=1}^{K}\frac{g_{nj}}{1+a_{nj}}} \tag{45}$$

we can write $a_{nk} = a_n g_{nk}$ and

$$-\frac{1}{L}\log_2 \det\left(\mathbf{X}^\dagger \mathbf{G}_n \mathbf{X}\right) \approx \log_2 e + \log_2\left(\frac{a_n}{L}\right) - \frac{1}{L}\sum_{k=1}^{K}\log_2(1+a_n g_{nk}) \tag{46}$$

with a_n the nonnegative solution to

$$\sum_{k=1}^{K}\frac{g_{nk}}{g_{nk}+1/a_n} = L. \tag{47}$$

C. Details of Examples 3 and 7

If the integers u and v denote the indices of a cell on the axes shown in the inset of Fig. 3, the cartesian coordinates of the corresponding BS are

$$x = \frac{3}{2}uR \tag{48}$$

$$y = \sqrt{3}\left(v + \frac{u}{2}\right)R. \tag{49}$$

Relative to its serving BS, the relative cartesian position of a user centered in azimuth in its sector and at distance $2R/3$ from the base is, for each of the sectors as labeled in Fig. 3,

$$\Delta x_1 = -\frac{R}{3} \qquad \Delta x_2 = -\frac{R}{3} \qquad \Delta x_3 = \frac{2R}{3} \tag{50}$$

$$\Delta y_1 = \frac{R}{\sqrt{3}} \qquad \Delta y_2 = -\frac{R}{\sqrt{3}} \qquad \Delta y_3 = 0. \tag{51}$$

Therefore, the distance between the BS at the origin ($u = v = 0$) and the users at the cell with indices u and v equals, for each of the sectors therein,

$$d_1(u, v) = R\sqrt{\left(\frac{3}{2}u - \frac{1}{3}\right)^2 + 3\left(v + \frac{u}{2} + \frac{1}{3}\right)^2} \tag{52}$$

$$d_2(u, v) = R\sqrt{\left(\frac{3}{2}u - \frac{1}{3}\right)^2 + 3\left(v + \frac{u}{2} - \frac{1}{3}\right)^2} \tag{53}$$

$$d_3(u, v) = R\sqrt{\left(\frac{3}{2}u + \frac{2}{3}\right)^2 + 3\left(v + \frac{u}{2}\right)^2}. \tag{54}$$

The corresponding normalized power gains are $g_1(u, v) = D d_1^{-\gamma}$, $g_2(u, v) = D d_2^{-\gamma}/Q$ and $g_3(u, v) = D d_3^{-\gamma}/Q$, with D the constant that renders

$$\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} (g_1(u, v) + g_2(u, v) + g_3(u, v)) = 1 \tag{55}$$

where Q is the antenna front-to-back ratio. (Because of symmetries, applying the factor $1/Q$ to sectors 2 and 3 of each cell is equivalent to applying it to all three sectors in the slices of the system spanned by sectors 2 and 3 of the central reference cell.) With $\gamma = 3.8$, (55) gives $D = 0.157R$. Then, for Example 3, we can rewrite (7) as

$$\mathsf{SIR}_n = \frac{\sum_{k=1}^{K} g_{nk}}{\sum_{k=K+1}^{\tilde{K}} g_{nk}} \tag{56}$$

and, using all the foregoing quantities, compute each SINR_n by having the appropriate terms in the numerator and denominator of (56) with $\tilde{K} \to \infty$.

For Example 7, a is the solution to

$$\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} \left(\frac{g_1(u, v)}{g_1(u, v) + 1/a} + \frac{g_2(u, v)}{g_2(u, v) + 1/a} + \frac{g_3(u, v)}{g_3(u, v) + 1/a}\right) = L \tag{57}$$

while

$$C_\infty = \log_2 \frac{ae}{L} - \frac{1}{L}\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} [\log_2(1 + a\, g_1(u, v)) + \log_2(1 + ag_2(u, v))$$

$$+ \log_2(1 + ag_3(u, v))] \tag{58}$$

Numerical evaluation of (57) and (58) with $Q = 100$ yields the results in Example 7.

REFERENCES

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927 – 1938, nov 2003.

[2] N. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. on Inform. Theory*, vol. 50, no. 12, pp. 3062– 3080, Dec. 2004.

[3] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 362 − 371, Feb. 2004.

[4] G. J. Foschini, K. Karakayali, and R. A. Valenzuela, "Enormous spectral efficiency of isolated multiple antenna links emerges in a coordinated cellular network," *IEE Proceedings - Communications*, vol. 153, no. 4, pp. 548–55, Aug. 2006.

[5] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Local base station cooperation via finite-capacity links for the uplink of linear cellular networks," *IEEE Trans. Info. Theory*, vol. 55, no. 1, pp. 190–204, Jan. 2009.

[6] Coordinated multi-point operation for LTE physical layer aspects (Release 11), *3GPP TR 36.819*, 2011.

[7] A. Gorokhov, "Coordinated joint transmission in WWAN," *IEEE Communication Theory Workshop (CTW'10), available at "http://www.ieee-ctw.org/2010/mon/Gorokhov.pdf"*, May 2010.

[8] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102 –111, feb 2011.

[9] A. Barbieri, P. Gaal, S. Geirhofer, T. Ji, D. Malladi, Y. Wei, and F. Xue, "Coordinate downlink multi-point communications in heterogeneous 4G cellular networks," in *Information Theory and Applications Workshop (ITA)*, Feb. 2012.

[10] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *To appear, IEEE Journal on Sel. Areas in Comm.*, Apr. 2012.

[11] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[12] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom for the K user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[13] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[14] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[15] V. R. Cadambe, K. Gomadam, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inform. Theory*, vol. 57, no. 6, pp. 3309–3322, June 2011.

[16] S. W. Peters and R. W. Heath, Jr., "Algorithms for the MIMO interference channel," *submitted to IEEE Transactions on Signal Processing*, July 2009.

[17] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.

[18] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and Wei Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE Journal on Sel. Areas in Communications*, vol. 28, no. 9, pp. 1380 –1408, Dec. 2010.

[19] P. Wang, H. Wang, L. Ping, and X. Lin, "On the capacity of MIMO cellular systems with base station cooperation," *IEEE Trans. on Wireless Communications*, vol. 10, no. 11, pp. 3720–3731, Nov. 2011.

[20] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. M. Zaidel, H. V. Poor, and S. Shamai, "Information theoretic considerations for wireless cellular systems: The impact of cooperation," *Foundations and Trends in Communications and Information Theory*, vol. 7, 2012.

[21] R. Tresch and M. Guillaud, "Clustered interference alignment in large cellular networks," in *Int'l Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC'09)*, Sept. 2009, pp. 1024–1028.

[22] W. Mennerich and W. Zirwas, "Reporting effort for cooperative systems applying interference floor shaping," in *Int'l Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC'11)*, Sept. 2011.

[23] C. Suh, M. Ho, and D. N. C. Tse, "Downlink interference alignment," *IEEE Trans. on Communications*, vol. 59, no. 9, pp. 2616–2626, Sept. 2011.

[24] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. on Inform. Theory*, vol. 54, no. 12, pp. 5534–5562, 2008.

[25] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian K user interference channel," *IEEE Trans. on Inform. Theory*, vol. 56, no. 7, pp. 3297–3303, 2010.

[26] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. on Communications*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.

[27] A. Lozano and D. Porrat, "Non-peaky signals in wideband fading channels: achievable bit rates and optimal bandwidth," *IEEE Trans. on Wireless Communications*, vol. 11, no. 1, pp. 246–257, Jan. 2012.

[28] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[29] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *Asilomar Conference on Signals, Systems and Computers*, 2007, pp. 83–87.

[30] S. Venkatesan, H. Huang, A. Lozano, and R. Valenzuela, "A WiMAX-based implementation of network MIMO for indoor wireless systems," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 9, Oct. 2009.

[31] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. on Wireless Communications*, vol. 8, no. 6, pp. 2983–2989, June 2009.

[32] S. A. Ramprashad, G. Caire, and H. C. Papadopoulos, "Cellular and network MIMO architectures: MU-MIMO spectral efficiency and costs of channel state information," in *Asilomar Conference on Signals, Systems and Computers*, Nov. 2009, pp. 1811–1818.

[33] S. W. Peters and R. W. Heath Jr., "User partitioning for less overhead in MIMO interference channels," *IEEE Trans. Wireless Communications (to appear), preprint at arXiv:1007.0512*, 2012.

[34] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," *Proc. of 37th Annual Allerton Conf. on Communication, Control and Computing, Monticello, IL*, vol. 37, pp. 958–966, Sept. 1999.

[35] A. Lapidoth and S. Shamai, "Fading channels: How perfect need 'perfect side information' be?," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[36] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. on Inform. Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.

[37] F. Rusek, A. Lozano, and N. Jindal, "Mutual information of IID complex Gaussian signals on block Rayleigh-faded channels," *IEEE Trans. Inform. Theory*, vol. 58, no. 1, Jan. 2012.

[38] A. Lozano, A. M. Tulino, and S. Verdu, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.

[39] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2491–2509, Aug. 2005.

[40] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing," in *IEEE Int'l Conf. on Communications (ICC'08)*, 2008, pp. 4033–4037.

[41] I. D. Garcia, N. Kusashima, K. Sakaguchi, and K. Araki, "Dynamic cooperation set clustering on base station cooperation cellular networks," in *IEEE Int'l Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC'10)*, 2010, pp. 2127–2132.

[42] L. Thiele, T. Wirth, K. Brner, M. Olbrich, V, Jungnickel, J. Rumold, and S. Fritze, "Modeling of 3D field patterns of downtilted antennas and their impact on cellular systems," *International ITG Workshop on Smart Antennas (WSA 2009)*, Feb. 2009.

[43] T.L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Communications*, vol. 9, no. 11, pp. 3590 –3600, Nov. 2010.

[44] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal Multivariate Analysis*, vol. 54, pp. 175–192, 1995.